R24

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/944,474	08/31/2001	Dan Gallivan	013.0171.01	2388

22895 7590 04/28/2004

PATRICK J S INOUYE P S
810 3RD AVENUE
SUITE 258
SEATTLE, WA 98104

EXAMINER	
LY, ANH	

ART UNIT	PAPER NUMBER
2172	3

DATE MAILED: 04/28/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 09/944,474	Applicant(s) GALLIVAN ET AL.	
	Examiner Anh Ly	Art Unit 2172	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *31 August 2001*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-44* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-44* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date #2.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. This Office Action is response to Applicants' communications filed on 08/31/2001.

2. Claims 1-44 are pending in this application.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to

consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

5. Claims 1-44 are rejected under 35 U.S.C. 103(a) as being unpatentable over US Patent No. 6,510,406 issued to Marchisio in view of US Patent No. 6,701305 issued to Holt et al. (hereinafter Holt).

With respect to claim 1, Marchisio teaches a histogram module determining a frequency of occurrences of concepts in a set of unstructured documents, each concept representing an element occurring in one or more of the unstructured documents (the frequency of occurrences of individual terms based on the extraction based on concepts: see fig. 1 and col. 45-65);

a selection module selecting, a subset of concepts out of the frequency of occurrences (parsing the user query into terms or phrases and the proximity based on the concepts: col. 7, lines 50-58),

grouping one or more concepts from the concepts subset (grouping the concepts: col. 15, lines 60-67 and col. 16, lines 1-8; also see fig. 7);

assigning weights to one or more clusters of concepts for each group of concepts (assigning weight to the term in the user query : col. 15, lines 1-16 and col. 2, lines 25-42); and

each document indexed by each such group of concepts between the frequency of occurrences and the weighted cluster (col. 15, lines 1-16 and col. 9, lines 52-67 and col. 10, lines 1-30).

Marchisio discloses searching or retrieving by latent concept or latent semantic for the fundamental problems of synonymy and polysemy in the text mining and using data mining techniques in order to overcome a wide margin of uncertainty in the initial choice of a keyword in a query, from which the user can query unstructured document such as electronic message or document) or structured document such as document storing in the database with indexing, clustering of documents with the concepts (col. 15, lines 1-16, see abstract, fig. 2) and indexing document (col. 9, lines 52-67 and col. 10, lines 1-30). Marchisio also teaches the computational approximation of query (fig. 6 and col. 15, lines 25-50). Marchisio does not explicitly teach a best-fit module calculating a best-fit approximation for each document for each such concept grouped into the group of concepts.

However Holt teaches approximating some of the semantics latent in the documents for the synonymy and polysemy in the documents (col. 3, lines 40-67 and col. 4, lines 1-40).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Marchisio with the teachings of Holt so as to have a approximation for capturing the semantics latent in each document (col. 3, lines 40-67). The motivation is that the ability of search engines interacts with the user query to be searched based on the concepts or semantic information and to list a list of relevant documents that do not contain the exact terms using in the user query (col. 5, lines 35-45).

With respect to claims 2-3, Marchisio teaches an extraction module extracting features from each of the unstructured documents and normalizing the extracted features into the concepts and a structured database storing the extracted features as uniquely identified records (see fig. 2, feature extraction from a relational database and col. 9, lines 8, lines 33-67 and col. 9, lines 1-15).

With respect to claim 4, Marchisio teaches a visualization module visualizing the frequency of occurrences, comprising at least one of creating a histogram mapping the frequency of occurrences for each document in the unstructured documents set and creating a corpus graph mapping the frequency of occurrence for all such documents in the unstructured documents set (see figs 8-9. col. 16, lines 33-58).

With respect to claim 5, Marchisio teaches a threshold comprising a median and edge conditions, each such concept in the concepts subset occurring within the edge conditions (col. 6, lines 55-65 and col. 7, lines 18-27 and fig. 1).

With respect to claims 6-7, Marchisio teaches an inner product module determining, for each group of concepts, the best fit approximation as the inner product between the frequency of occurrences and the weighted cluster for each such concept in the group of concepts, and wherein the inner product d (cluster) is calculated according to the equation comprising: d(cluster) = ~ doc(termi) * cluster (termi) where doc(concept) represents the frequency of occurrence for a given concept in the document and cluster(concept) represents the weight for a given cluster (inner product's computation and its equation: col. 2, lines 3-42 and approximation for the query: fig. 6 and col. 15, lines 25-50).

With respect to claim 8, Marchisio teaches a system as discussed in claim 1.

Marchisio discloses searching or retrieving by latent concept or latent semantic for the fundamental problems of synonymy and polysemy in the text mining and using data mining techniques in order to overcome a wide margin of uncertainty in the initial choice of a keyword in a query, from which the user can query unstructured document such as electronic message or document) or structured document such as document storing in the database with indexing, clustering of documents with the concepts (col. 15, lines 1-16, see abstract, fig. 2) and indexing document (col. 9, lines 52-67 and col. 10, lines 1-30). Marchisio also teaches the computational approximation of query (fig. 6 and col. 15, lines 25-50). Marchisio does not explicitly teach a control module iteratively re-determining the best-fit approximation to a change in the set of unstructured documents.

However Holt teaches approximating some of the semantics latent in the documents for the synonymy and polysemy in the documents (col. 3, lines 40-67 and col. 4, lines 1-40).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Marchisio with the teachings of Holt so as to have a approximation for capturing the semantics latent in each document (col. 3, lines 40-67). The motivation is that the ability of search engines interacts with the user query to be searched based on the concepts or semantic information and to list a list of relevant documents that do not contain the exact terms using in the user query (col. 5, lines 35-45).

Claim 9 is essentially the same as claim 1 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 1 hereinabove.

Claim 10 is essentially the same as claim 2 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 2 hereinabove.

Claim 11 is essentially the same as claim 3 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 3 hereinabove.

Claim 12 is essentially the same as claim 4 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 4 hereinabove.

Claim 13 is essentially the same as claim 5 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 5 hereinabove.

Claim 14 is essentially the same as claim 6 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 6 hereinabove.

Claim 15 is essentially the same as claim 7 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 7 hereinabove.

Claim 16 is essentially the same as claim 8 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 8 hereinabove.

Claim 17 is essentially the same as claims, 9, 10, 11, 12, 13, 14, 15 or 16 except that it is directed to a computer-readable storage medium rather than a method, and is rejected for the same reason as applied to the claims 9, 10, 11, 12, 13, 14, 15 or 16 hereinabove.

With respect to claim 18, Marchisio teaches an extraction module extracting a multiplicity of concepts from a set of unstructured documents into a lexicon uniquely identifying each concept and a frequency of occurrence (see fig. 2, item 21 and 28, and also see fig. 1, item 16, col. 7, lines 28-38": parsing the unstructured text into lexical uniquely identifier for each concept (concept ID number); the frequency of occurrences of individual terms based on the extraction based on concepts: see fig. 1 and col. 45-65);

a frequency-mapping module creating a frequency of occurrence representation for each documents set, the representation providing an ordered corpus of the frequencies of occurrence of each concept (best match model: col. 2, lines 3-56 and col. 8, lines 20-32);

a concept selection module selecting a subset of concepts from the frequency of occurrence representation filtered against a minimal set of concepts each referenced in at least two documents with no document in the corpus being unreferenced (parsing the user query into terms or phrases and the proximity based on the concepts: col. 7, lines

50-58; a set of documents to be inspected and some not: to be searched or extracted: col. 11, lines 20-46);

a group generation module generating a group of weighted clusters of concepts selected from the concepts subset (grouping the concepts: col. 15, lines 60-67 and col. 16, lines 1-8; also see fig. 7); and

each document weighted against each group of weighted clusters of concepts (col. 15, lines 1-16 and col. 9, lines 52-67 and col. 10, lines 1-30)

Marchisio discloses searching or retrieving by latent concept or latent semantic for the fundamental problems of synonymy and polysemy in the text mining and using data mining techniques in order to overcome a wide margin of uncertainty in the initial choice of a keyword in a query, from which the user can query unstructured document such as electronic message or document) or structured document such as document storing in the database with indexing, clustering of documents with the concepts (col. 15, lines 1-16, see abstract, fig. 2) and indexing document (col. 9, lines 52-67 and col. 10, lines 1-30). Marchisio also teaches the computational approximation of query (fig. 6 and col. 15, lines 25-50). Marchisio does not explicitly teach a best-fit module calculating a best-fit approximation for each document for each such concept grouped into the group of concepts.

However Holt teaches approximating some of the semantics latent in the documents for the synonymy and polysemy in the documents (col. 3, lines 40-67 and col. 4, lines 1-40).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Marchisio with the teachings of Holt so as to have a approximation for capturing the semantics latent in each document (col. 3, lines 40-67). The motivation is that the ability of search engines interacts with the user query to be searched based on the concepts or semantic information and to list a list of relevant documents that do not contain the exact terms using in the user query (col. 5, lines 35-45).

With respect to claims 19-22, Marchisio teaches a histogram module creating a histogram mapping the frequency of occurrence representation for each document in the documents set (generating the term-document matrix to indicating the number of occurrences of the term: see fig. 1, item 6, col. 6, lines 35-65; the frequency of occurrences of individual terms based on the extraction based on concepts: see fig. 1 and col. 45-65);

a data mining module mining the multiplicity of concepts from each document as at least one of a noun, noun phrase and tri-gram (col. 2, lines 57-67);

a normalizing module normalizing the multiplicity of concepts into a substantially uniform lexicon (10, 10-31); and

wherein the substantially uniform lexicon is in third normal form (col. 12, lines 1-36).

With respect to claim 23, Marchisio teaches a corpus mapping module creating a corpus graph mapping the frequency of occurrence representation for all documents in the documents set (see figs 8-9. col. 16, lines 33-58).

With respect to claim 24, Marchisio teaches a threshold module defining the pre-defined threshold as a median value and a set of edge conditions and choosing those concepts falling within the edge conditions as the concepts subset (col. 6, lines 55-65 and col. 7, lines 18-27 and fig. 1).

With respect to claim 25-26, Marchisio teaches a cluster module naming, one or more of the concepts within the concepts subset to a cluster and assigning a weight to each concept with each such cluster (assigning weight to the term in the user query : col. 15, lines 1-16 and col. 2, lines 25-42; filtering and identifying the cluster: col. 3, lines 5-12); and

.a group module grouping, one or more of the clusters into each such group of weighted clusters of concepts (assigning weight to the term in the user query : col. 15, lines 1-16 and col. 2, lines 25-42 and partitioning or grouping of documents: col. 15, lines 40-50).

With respect to claim 27, Marchisio teaches a system as discussed in claim 18. Marchisio discloses searching or retrieving by latent concept or latent semantic for the fundamental problems of synonymy and polysemy in the text mining and using data mining techniques in order to overcome a wide margin of uncertainty in the initial choice of a keyword in a query, from which the user can query unstructured document such as electronic message or document) or structured document such as document storing in the database with indexing, clustering of documents with the concepts (col. 15, lines 1-16, see abstract, fig. 2) and indexing document (col. 9, lines 52-67 and col. 10, lines 1-30). Marchisio also teaches the computational approximation of query (fig. 6 and col.

15, lines 25-50) and building the computation of the distance between the query vector and document clusters in the optimization problem. Marchisio does not explicitly teach a Euclidean module calculating a Euclidean distance between the frequency of occurrence for each document and a corresponding weighted cluster.

However Holt teaches the Euclidean distance of the vector (col. 4, lines 25-48).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Marchisio with the teachings of Holt so as to have a way to computing the distance query vector and document cluster (col. 4, lines 25-48). The motivation is that the ability of search engines interacts with the user query to be searched based on the concepts or semantic information and to list a list of relevant documents that do not contain the exact terms using in the user query (col. 5, lines 35-45).

With respect to claim 28, Marchisio teaches a iteration module removing select documents from the documents set and iteratively reevaluating the matrix of best fit approximations based on a revised frequency of occurrence representation and concepts subset (col. 14, lines 56-67; and removing the documents: col. 7, lines 55-65 and col.4, lines 30-36).

With respect to claims 29-30, Marchisio teaches a structured database storing the lexicon, the lexicon comprising a plurality of records each uniquely identifying one such concept and an associated frequency of occurrence (see fig. 1 and fig 2, col. 9, lines 8-14); and

wherein the structured database is an SOL database (col. 10, lines 32-45 and lines 58-67 and col. 11, lines 1-5, see fig. 2).

Claim 31 is essentially the same as claim 18 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 18 hereinabove.

Claim 32 is essentially the same as claim 19 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 19 hereinabove.

Claim 33 is essentially the same as claim 20 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 20 hereinabove.

Claim 34 is essentially the same as claim 21 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 21 hereinabove.

Claim 35 is essentially the same as claim 22 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 22 hereinabove.

Claim 36 is essentially the same as claim 23 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 23 hereinabove.

Claim 37 is essentially the same as claim 24 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 24 hereinabove.

Claim 38 is essentially the same as claim 18 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 25 hereinabove.

Claim 39 is essentially the same as claim 18 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 26 hereinabove.

Claim 40 is essentially the same as claim 18 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 27 hereinabove.

Claim 41 is essentially the same as claim 18 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 28 hereinabove.

Claim 42 is essentially the same as claim 18 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 29 hereinabove.

Claim 43 is essentially the same as claim 18 except that it is directed to a method rather than a system, and is rejected for the same reason as applied to the claim 30 hereinabove.

Claim 44 is essentially the same as claims, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41 or 42 except that it is directed to a computer-readable storage medium rather than a method, and is rejected for the same reason as applied to the claims, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41 or 42 hereinabove.

Contact Information

6. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Anh Ly whose telephone number is 703 306-4527 or via E-Mail: ANH.LY@USPTO.GOV. The examiner can normally be reached on 7:30 AM - 4:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Breene, can be reached on 703 305-9790. The fax phone number for the organization where this application or proceeding is assigned is 703 746-7239.

Any response to this action should be mailed to:

Commissioner of Patents and Trademarks

Washington, D.C. 20231

or faxed to: Central Office (703) 872-9306 (Central Official Fax Number)

Hand-delivered responses should be brought to Crystal Park II, 2121 Crystal Drive, Arlington, VA, Fourth Floor (receptionist).

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703 308-6606 or 703 305-3900.

JEAN M. CORRIELUS
PRIMARY EXAMINER

ANH LY
APR. 23rd, 2004